ERASMUS

ECONOMETRIC INSTITUTE

POSTERIOR MOMENTS COMPUTED BY MIXED INTEGRATION

H.K. VAN DIJK, T. KLOEK AND C.G.E. BOENDER

REPORT 8520/A

ERASMUS UNIVERSITY ROTTERDAM - P.O. BOX 1738 - 3000 DR ROTTERDAM - THE NETHERLANDS

POSTERIOR MOMENTS COMPUTED BY MIXED INTEGRATION*

H.K. van Dijk**
Erasmus University Rotterdam/National Bureau of
Economic Research, Cambridge, MA 02138, U.S.A.

T. Kloek
Erasmus University Rotterdam, Rotterdam, The Netherlands

C.G.E. Boender
Erasmus University Rotterdam, Rotterdam, The Netherlands
Delft University of Technology, Delft, The Netherlands

Abstract

A flexible numerical integration method is proposed for the computation of moments of a multivariate posterior density with different tail properties in different directions. The method (called mixed integration) amounts to a combination of classical numerical integration and Monte Carlo integration. Mixed integration is parsimonious in the sense that it makes use of the same parameters as the more restrictive multivariate normal importance function. The method is applied in order to compute the posterior scores of three candidates for a professorship in Operations Research, taking into account four different decision criteria.

* This is a revision and extension of Report 8318/E. We are indebted to a referee for some helpful comments.
** The research described in this paper was partially supported by NATO Science Fellowship no. N46-160. The final version of this paper was finished, while the first author was visiting Harvard University and the NBER. Their hospitality is gratefully acknowledged.

Contents

February 1985

1. INTRODUCTION

Our research is directed towards the efficient computation of posterior moments of measurable functions of parameters of econometric models in multivariate cases.

Given our prior assumptions these posterior moments cannot be evaluated using analytical integration methods without gross approximation errors or excessive computational costs. In earlier work [4, 8] we made use of Monte Carlo integration methods [2, 6]. As an example of an econometric model we took the well known simultaneous equation model, which is nonlinear in the sense that the expected values of the endogenous variables are usually nonlinear functions of the parameters of interest θ. As a Monte Carlo integration method we made use of importance sampling.

Our application of importance sampling may be described briefly as follows. Let $g(\theta)$ be a measurable function of the parameters of interest θ, where g may be a scalar, a vector or a matrix. Standard examples of g are elements of θ and $\theta\theta'$, but we are also interested in marginal posterior densities [4] and in nonlinear functions of θ such as short-run and long-run multipliers [8]. Our purpose is the efficient computation of

$$(1) \qquad E\, g(\theta) = \frac{\int g(\theta)w(\theta)I(\theta)d\theta}{\int w(\theta)I(\theta)d\theta}$$

The weight function $w(\theta)$ is defined as $p(\theta)/I(\theta)$, where $p(\theta)$ is a kernel of a multivariate posterior density function and the importance function $I(\theta)$ is a multivariate density function with properties to be discussed below. The region of integration in (1) is the set of all θ satisfying $I(\theta) > 0$. The importance function is supposed to be a good approximation of $p(\theta)$, so that $w(\theta)$ is roughly a constant, and the importance function is supposed to have convenient Monte Carlo properties, so that it is relatively easy to generate

random drawings from it. As an importance function we made use of the truncated multivariate Student t density. For details we refer to [4, 8]. We shall use the term simple importance sampling for the approach used in [4, 8].

In several econometric applications we found that the surface of the posterior density of a nonlinear model may be ill behaved. An important reason for this phenomenon is small sample size. Then the question arises whether such a density can be approximated with a reasonable degree of accuracy by a multivariate Student t density. A limitation of the Student t is its symmetry, while truncation of the density is helpful only in special cases. Since our experience with the Student t density was not always successful, we started to consider alternative approaches.

In the present paper we consider the case where the kernel of the posterior density $p(\theta)$ is unimodal but has different tail properties in different directions and we propose a flexible numerical integration method, which is intended to handle this situation. The method proposed transforms the s-dimensional parameter space of vectors θ into another s-dimensional parameter space of pairs (η, ρ), where η is an $(s-1)$-dimensional vector and ρ a scalar. For ρ we take $\pm d$, where d is a measure of the distance between a point $\theta^{(i)}$, generated at random, and a point θ^0, a location estimate of θ. As a location estimate we take e.g. the posterior mode or a preliminary estimate of the posterior mean. The vector η is taken as the direction $(\theta^{(i)} - \theta^0)/\rho^{(i)}$ with one coordinate deleted to avoid degeneracy. After having performed the transformation, we generate a vector $\eta^{(i)}$ by means of Monte Carlo and apply classical numerical integration with respect to ρ given $\eta^{(i)}$. So the method amounts to a combination of classical numerical integration and Monte Carlo integration. We call it the <u>mixed integration</u> method. More details on the transformation are given in Section 2 and the Appendix B. The mixed integration method is discussed in Section 3. Further comments are contained in Section 4, while our conclusions are given in

Section 5. An application of mixed integration is presented in Appendix A.

2. A TRANSFORMATION

In this section we describe a particular transformation of a multivariate normal random variable. We start by defining a multivariate normal density function on the parameter space as

$$(2) \qquad I(\theta) = (2\pi)^{-\frac{1}{2}s}|V|^{-\frac{1}{2}}\exp[-\tfrac{1}{2}(\theta - \theta^0)'V^{-1}(\theta - \theta^0)]$$

with $-\infty < \theta_j < \infty$, $j = 1, \ldots, s$. Here θ^0 is the posterior mode or a (possibly rough) first stage estimate of the posterior mean; V is minus the inverse of the Hessian of the log posterior evaluated at θ^0 or another preliminary estimate of the posterior covariance matrix. For more details on the choice of parameters for importance functions we refer to [8, Section 4; and 10].

Next, we partition $\theta - \theta^0$ and V^{-1} as

$$(3) \qquad \theta - \theta^0 = \begin{bmatrix} u \\ v \end{bmatrix} \qquad V^{-1} = \begin{bmatrix} P & q \\ q' & r \end{bmatrix}$$

where v and r are scalars.

We shall make use of

$$(4) \qquad d := (u'Pu + 2u'qv + rv^2)^{\frac{1}{2}}$$

$$(5) \qquad \bar{v} := E(v|u) = -\frac{u'q}{r}$$

Note that d measures the distance from θ to θ^0 according to a norm based on V^{-1} and that $\bar{v}$ is the mean value of the conditional normal distribution of v given u, which is called the regression function [1, p.29]. We partition the

region of integration of θ as

$$(6) \qquad \begin{aligned} S_1 &= \{\theta \mid \theta \in \mathbb{R}^s \text{ and } v \geq \bar{v}\} \\ S_2 &= \{\theta \mid \theta \in \mathbb{R}^s \text{ and } v < \bar{v}\} \end{aligned}$$

The transformation that carries θ into (η, ρ) is given by the following transformation formulae,

$$(7) \qquad \begin{aligned} (\eta, \rho) &= T_1(\theta) = \left(\frac{u}{d}, d\right) \quad \text{if } v \geq \bar{v} \\ (\eta, \rho) &= T_2(\theta) = \left(-\frac{u}{d}, -d\right) \quad \text{if } v < \bar{v} \end{aligned}$$

The regions of integration of (η, ρ) are given by

$$(8) \qquad \begin{aligned} S_1' &= \{(\eta, \rho) \mid \rho \in \mathbb{R}^+, \eta \in \Omega\} \\ S_2' &= \{(\eta, \rho) \mid \rho \in \mathbb{R}^-, \eta \in \Omega\} \end{aligned}$$

where $\mathbb{R}^+$ and $\mathbb{R}^-$ are the sets of positive and negative real numbers, respectively. The characterization of Ω will be discussed in the Appendix. Note that the sign of ρ determines the region of integration S_1' or S_2'.

The inverse transformation that carries (η, ρ) into θ is defined on S_1' and S_2' as the function

$$(9) \qquad (u, v) = T^{-1}(\eta, \rho) = \left(\rho\eta, -\frac{q'\eta}{r}\rho + \frac{\rho}{r}\sqrt{D(\eta)}\right)$$

with

$$D(\eta) = (q'\eta)^2 + r(1 - \eta'P\eta) \tag{10}$$

where we make use of (4) and (7). Note that v can be rewritten as

$$v = \bar{v} + \frac{\rho}{r}\sqrt{D(\eta)} \tag{11}$$

So the transformation $T_i(\theta)$ maps the elements of S_i onto S_i', i = 1, 2. The purpose of the inequality conditions in (7) is to make this transformation one-to-one. We shall give a geometric interpretation below.

The Jacobian determinant J, obtained by taking derivatives of u and v with respect to η and ρ, equals

$$J = \rho^{s-1}J_2(\eta) \tag{12}$$

with

$$J_2(\eta) = D^{-\frac{1}{2}}(\eta) \tag{13}$$

which can straightforwardly be verified. So the absolute value of the Jacobian determinant $|J|$ is the same in both regions of integrations.

The density function of (η, ρ) is given by

$$\begin{aligned} I^*(\eta, \rho) &= I[T^{-1}(\eta, \rho)]|J| \\ &= (2\pi)^{-\frac{1}{2}s}|V|^{-\frac{1}{2}}\exp(-\tfrac{1}{2}\rho^2)|\rho|^{s-1}J_2(\eta) \end{aligned} \tag{14}$$

where we have deleted an index for the region of integration S_1' or S_2' [5,

pp.156 and 157], because the density of (η, ρ) is invariant with respect to changing the region of integration. This result is due to the symmetry properties of the multivariate normal density.

A geometric interpretation of the transformation of the random variables θ is presented in Figure 1. At the top we have drawn a contour of a bivariate

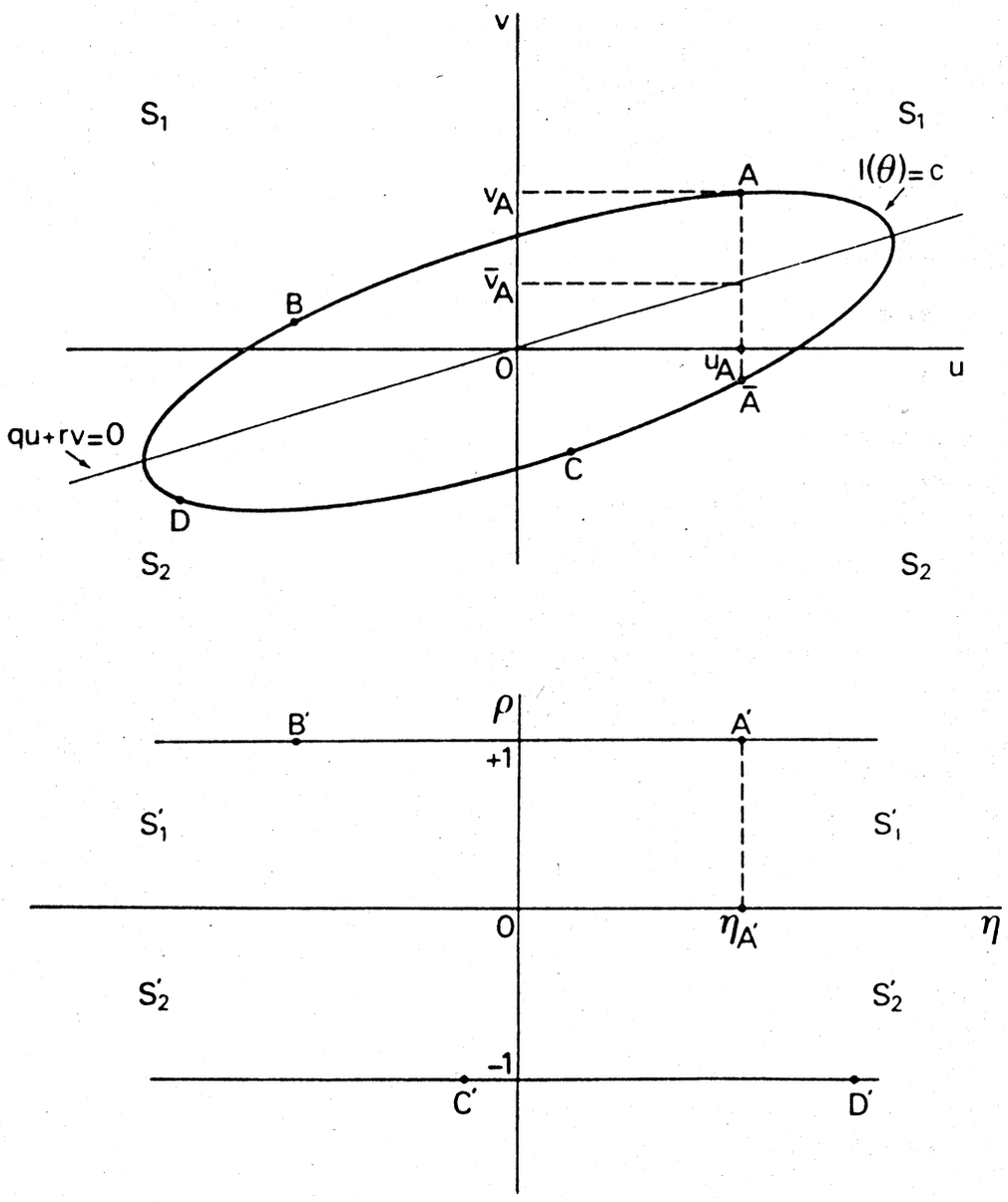

Fig.1 A transformation of the bivariate normal random variable $\theta=(u,v)$ into (η,ρ)

EGI 113/83

normal importance function $I(\theta) = c$ with $\theta^0 = 0$ and $d = 1$, for convenience only. The regression line $qu + rv = 0$ partitions the region of integration of θ-elements into S_1 and S_2 [compare (6)]. Consider the point A generated at random from $I(\theta)$. A is an element of S_1 because $v_A > \bar{v}_A$ [see (7)]. It follows that A is mapped into the point $A' \in S_1'$ at the bottom of Figure 1, where A' has the coordinates $\eta_{A'} = u_A$ and $\rho_{A'} = +1$. A similar analysis can be given for B. The points C and D are elements of S_2 for which $\rho = -1$ and $\eta = -u$. Note that C' and D' [in the (η, ρ) space] are reflected with respect to the vertical axis compared with C and D.

We conclude this section with a remark. If we do not use the inequality conditions of (7), the transformation of θ into (η, ρ) will be two-to-one in the following way. Let $\eta = u/\rho$ and $\rho = d$, instead of the transformation (7). Then the inverse transformation $\theta = T^{-1}(\eta, \rho)$ will give two solutions of v that are not symmetric around θ^0. This is illustrated in Figure 1. Consider the point $A' = (\eta_{A'}, +1)$ at the bottom. Given $u_A = \eta_{A'}$ [compare (9)], it is seen that there exist two solutions of θ: the points A and $\bar{A}$ on the contour $I(\theta) = c$. This is a disadvantage for practical applications. Another disadvantage of the two-to-one transformation is the asymmetric treatment of the elements of θ. In order to define the direction η one may delete any element of the s-vector $(\theta - \theta^0)$ instead of the last one. In such a case one obtains s+1 solutions of θ by inverting the transformation. Because of these disadvantages we have constructed the transformation described.

3. THE MIXED INTEGRATION METHOD

In this section we apply the transformation of Section 2 to the integrals of equation (1). We write the numerator as $A = A_1 + A_2$ where

$$A_i = \int_{S_i} g(\theta)w(\theta)I(\theta)d\theta \qquad (i = 1, 2) \tag{15}$$

The denominator is handled in the same way after substitution of $g(\theta) = 1$. By making use of the inverse transformation (9) we can write A_i in terms of η and ρ. Let

$$g(\theta) = g[T^{-1}(\eta, \rho)] =: g^*(\eta, \rho) \tag{16}$$

$$w(\theta) = \frac{p[T^{-1}(\eta, \rho)]}{I[T^{-1}(\eta, \rho)]} =: w^*(\eta, \rho)$$

Using (12), (14) and (16) we find

$$w^*(\eta, \rho)I^*(\eta, \rho) = \frac{p[T^{-1}(\eta, \rho)]}{I[T^{-1}(\eta, \rho)]} I[T^{-1}(\eta, \rho)]|J| \tag{17}$$

$$= p[T^{-1}(\eta, \rho)]|\rho|^{s-1}J_2(\eta)$$

Note that the importance function has been dropped from (17). We shall comment on this result in the next section. By making use of (14) and (16) we can write the first of the two integrals in (15) as

$$A_1 = \iint_{S_1'} g^*(\eta, \rho)w^*(\eta, \rho)I^*(\eta, \rho)d\eta d\rho \tag{18}$$

We assume that appropriate regularity conditions are satisfied and substitute (17) into (18). Then one obtains

$$A_1 = \int_{\Omega} \{\int_{\mathbb{R}^+} g^*(\eta, \rho)p[T^{-1}(\eta, \rho)]|\rho|^{s-1}d\rho\}J_2(\eta)d\eta \tag{19}$$

Similar results can be obtained for A_2, where the integration with respect to ρ is over $\bar{\mathbb{R}}$. We recall from Section 2 that the importance function (14) is invariant with respect to the sign of ρ, but we emphasize that the expressions in (16)-(19) will usually not be invariant.

Now suppose we draw N vectors $\eta^{(i)}$ from a distribution with a density proportional to $J_2(\eta)$. Then we may estimate A_1 by means of

$$\bar{A}_1 \propto \frac{1}{N} \sum_{i=1}^{N} \{\int_{\mathbb{R}^+} g^*(\eta^{(i)}, \rho) p[T^{-1}(\eta^{(i)}, \rho)] |\rho|^{s-1}\} d\rho \tag{20}$$

The proportionality sign in (20) is necessary since we have not bothered about the question which part of the integration constant of (2) corresponds to η and which part to ρ. If we treat the numerator and the denominator of (1) in the same way the constants will cancel. The generation of the random drawings η is done by drawing θ from (2) and deriving η through (4) and (7). The one-dimensional integrals in (20) are calculated by means of classical numerical integration methods. There are two reasons for this choice. First, classical integration procedures are known to be quite efficient in handling one-dimensional integration problems. Second, in the type of nonlinear problems we are interested in, it will be difficult to find a good importance function, since for certain values of η the expression $p[T^{-1}(\eta, \rho)]$ considered as a function of ρ will decrease faster than a normal density as $|\rho|$ tends to infinity, while for other values of η it may decrease slowly as a Student t density with a few degrees of freedom only. Anyway, this is our experience with the examples we studied in [9, 10, 11]. In fact, the integrals in (20) may not converge at all for a diffuse prior when ρ tends to infinity. In our own applications we have taken a prior which is positive on a bounded region only, but the approach of this section can be applied to integration over $\mathbb{R}^s$ provided the integrals exist.

For the practical application of the integration approach we shall discuss one further development. Suppose $\theta^{(i)}$ is a point generated at random from the normal importance function (2). Given $\theta^{(i)}$ we know which transformation formula, $T_1(\theta)$ or $T_2(\theta)$, we have to take from (7). Suppose we take T_1 which gives $(\eta^{(i)}, \rho^{(i)})$. Numerical integration is then performed on the half-line through the points θ^0 and $\theta^{(i)}$, consisting of all positive multiples of $(\theta^{(i)} - \theta^0)$. In case one has to take T_2 (instead of T_1) one performs numerical integration along the half line consisting of all negative multiples of $(\theta^{(i)} - \theta^0)$. One may generate another direction $\eta^{(i+1)}$ (through $\theta^{(i+1)}$) as a next step. We recommend however a modification in the practical application of the integration procedure. Clearly the points $\theta^{(i)} - \theta^0$ and $\theta^0 - (\theta^{(i)} - \theta^0)$ are symmetric around θ^0. Suppose as before that $\theta^{(i)}$ is a point generated at random. Let $\theta^0 - (\theta^{(i)} - \theta^0)$ be the next point to be considered. Then one can perform one-dimensional numerical integration along the line consisting of θ^0 plus all positive and negative multiples of $(\theta^{(i)} - \theta^0)$. This method of generating one point by Monte Carlo and the next one by a symmetry argument is well known in the literature as <u>antithetic</u> sampling [2, 6]. It implies that we replace (20) by

$$(21) \qquad \hat{A} \propto \frac{1}{2N} \sum_{i=1}^{N} \{ \int_{-\infty}^{\infty} g^*(\eta^{(i)}, \rho) p[T^{-1}(\eta^{(i)}, \rho)] |\rho|^{s-1} d\rho \}$$

We divide by 2N since each point drawn is used twice. The posterior kernel is not defined at $\rho = 0$ but since this event has measure zero it is of no importance in the computation of the integrals.

In our integration approach the proportionality (21) is the basic formula for the estimation of the integral A. Since this approach amounts to a combination of classical numerical integration and Monte Carlo integration, we call it the <u>mixed integration</u> method. In the next section we give some further

comments.

4. FURTHER COMMENTS

First, it seems useful to compare the present approach with an alternative we tried before. In [10] we gave pictures of some very skew posterior densities. Student t densities necessarily yield bad approximations in such cases since they are essentially symmetric. Existing families of skew multivariate densities, such as Wishart and poly-t, might be considered but it is rather difficult to fit them to a given posterior. We also considered products of univariate skew densities fitted along the main axes, but the results were not encouraging. Compared with these alternatives the present approach has the advantage that it is both flexible and parsimonious. By flexible we mean that we perfectly follow the shape of the posterior when we integrate with respect to ρ. By parsimonious that we do not introduce additional parameters to describe the shape of the importance function, given preliminary posterior mode estimates or first-stage integration results on posterior mean and covariance matrix.

Second, one might observe that the reciprocal of the density function $I[T^{-1}(\eta, \rho)]$ does not appear in (19)-(21) in the same way it appears in (1) through $w(\theta)$. If we are interested in computing the integral

$$\int g(\theta)p(\theta)d\theta \tag{22}$$

and if $p(\theta)$ has no convenient Monte Carlo properties, we substitute $p(\theta)$ by $[p(\theta)/I(\theta)]I(\theta)$ in order to obtain an expression containing a density from which we can conveniently draw points $\theta^{(i)}$ at random. If we apply the transformation defined by (3)-(7), we can rewrite (22) as follows

$$(23) \qquad \int\int g[T^{-1}(\eta, \rho)]p[T^{-1}(\theta, \rho)]|J|d\rho d\eta =$$
$$\int\{\int g[T^{-1}(\eta, \rho)]p[T^{-1}(\eta, \rho)]|\rho|^{s-1}d\rho\}J_2(\eta)d\eta$$

compare Section 2 and, in particular, (12). Since $J_2(\eta)$ is already proportional to a density with convenient Monte Carlo properties there is no reason to introduce an importance function in this stage. In fact, when defining the transformation we have already implicitly introduced our importance function. Note that the posterior kernel of (η, ρ) is decomposed as a conditional density of ρ given η and a marginal density of η. We have incorporated the importance function $I(\theta)$ of (2) explicitly in our derivation since we want to emphasize the link between the kernel $J_2(\eta)$ and the normal density $I(\theta)$ that is used to generate the random directions $\eta^{(i)}$ $(i = 1, \ldots, N)$. We note that there exists a computational efficiency problem with respect to the generation of the random directions. This is illustrated in Figure 2a below [compare comment four].

Third, we mention that one can use the same integral with respect to ρ for different functions $g^*(\eta, \rho)$. For instance, when one is interested in the computation of the first order moments of the vector $\theta - \theta^0$ one takes for g the elements of u and v. But (given η) these are constant multiples of p [compare (9)]. As a consequence only one integral with respect to ρ has to be computed for all elements of θ in (21). So moments of linear functions of θ are easy to compute. This holds similarly for the zero order and second order moments of θ. Marginal posterior densities may be approximated by making use of the integration results for the case $g(\theta) = 1$ on some interval and zero elsewhere [4, Section 7]. Moments of complicated nonlinear functions of θ require additional numerical integrations.

Fourth, a geometric interpretation of some of the features of mixed integration is illustrated for $s = 2$ in Figures 2a and 2b. In Figure 2a we

have drawn a contour of a posterior density $p(\theta) = c$ with the mode $\theta^0 = 0$, for convenience only. The point A, generated at random from a bivariate normal importance function with $\theta^0 = 0$, is located on the contour $I(\theta) = c$. Numerical integration can be performed on the line through the points -A and A by making use of antithetic sampling. Note that (-)A is mapped into (-)A' [compare also Figure 1]. Typical shapes of the function $p[T^{-1}(\eta_{A'}, \rho)]|\rho|$, considered as a function of ρ given $\eta_{A'}$, are shown in Figure 2b. The different shapes illustrate the flexibility of mixed integration. Furthermore, Figure 2a illustrates that it is less efficient to take a standard normal importance function (V = I) because one should generate more lines in the direction of the main diagonal than orthogonal to it.

Finally, we mention two cases in the literature where a transformation of the multivariate normal density is considered with a Jacobian factor that differs from (12) and (13) only with respect to the expression of $J_2(\eta)$. First, Anderson [1, pp. 175 and 176] and Kendall and Stuart [3, pp. 246 and 247] give the transformation of normal variables into polar coordinates.

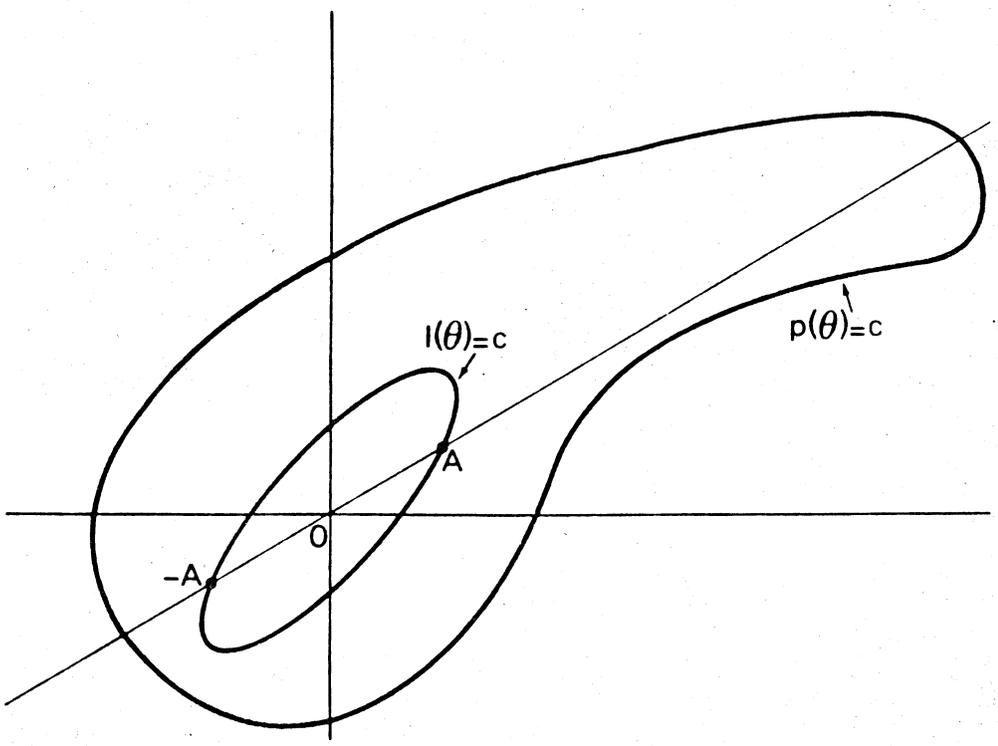

Fig. 2a Contours of the posterior density $p(\theta)=c$ and the normal importance function $I(\theta)=c$ with $\theta^{\circ}=0$

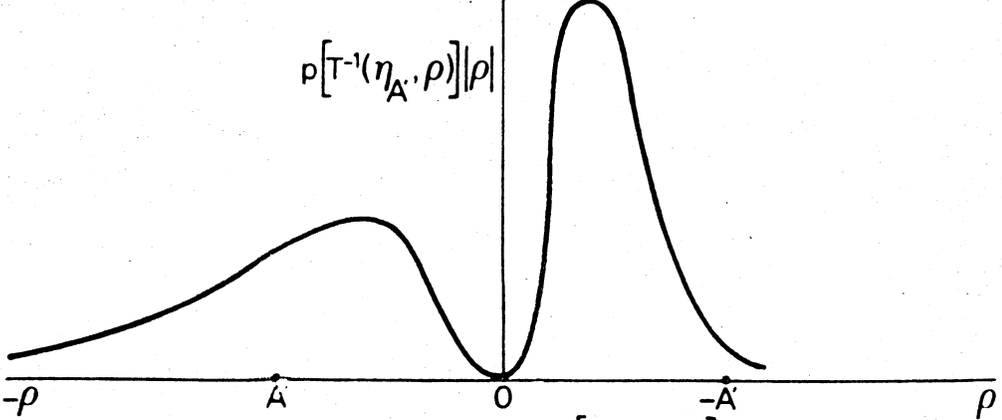

Fig. 2b Typical shapes of the function $p\left[T^{-1}(\eta_{A'}, \rho)\right]|\rho|$ considered as a function of ρ given $\eta_{A'}$

EGI 114/83

We have opted for (7) since it appears relatively easy for practical applications. Second, Trotter and Tukey [7] make use of a transformation of normal random variables in what is known as Conditional Monte Carlo. In this approach a possibly awkward parameter space is enlarged to a more suitably chosen space. Details are discussed in [2, Chapter 6]. By contrast, we have reduced the s-dimensional parameter space of θ-elements into an (s-1)-dimensional space of η-elements. Our approach has the advantage of being parsimonious with respect to the number of parameters of the importance function.

5. CONCLUSIONS

In this paper we have proposed a flexible numerical integration method that can be used for the computation of posterior moments in case the multivariate posterior density has different tail properties in different directions. The method is parsimonious because one makes only use of the parameters of the more restrictive multivariate normal importance function, which will approximate the tail behavior of the posterior mentioned above very poorly. Increased flexibility is, of course, not a free good. But the price of one-dimensional numerical integration along lines in the parameter space is not very restrictive on modern computers. Practical experience with several examples, for instance with the Klein-Goldberger model, which involves thirty-dimensional numerical integration [9], indicates the feasibility of the mixed integration approach in a case where simple importance sampling failed to converge. In [11] we compared mixed integration and simple importance sampling with an alternative Monte Carlo method where the importance function consists of a finite mixture of multivariate normal densities. The particular mixture used was tailor-made for the example studied. In other words, the finite

mixture approach can probably not be applied to arbitrary models, without first studying the properties of their likelihood functions. The results of some experiments indicate that an importance function based on mixtures is rather efficient but that mixed integration appears robust and can be used in a rather mechanical way. Of course, more experience is needed in this area before any final conclusions can be given.

APPENDIX A*

In this appendix we apply the mixed integration method in order to compute the scores $\omega_1, \ldots, \omega_n$ of n alternatives $A_1, \ldots, A_n$ ($0 \leq \omega_j \leq 1 \; \forall \; j = 1, \ldots, n; \; \Sigma_{j=1}^{n} \omega_j = 1$). The scores of the alternatives are obtained taking into account m different decision-criteria $C_1, \ldots, C_m$. More details are given below. As a particular example we consider the case presented in Lootsma (1980) where three candidates, A_1, A_2 and A_3, have applied for a professorship in Operations Research. In order to compare the abilities of the candidates an advisory committee of N=3 members has identified four decision-criteria $C_1, \ldots, C_4$. These criteria are mathematical creativity (C_1), creativity in implementations (C_2), administrative capabilities (C_3), and human maturity (C_4): the candidates remain anonymous.

A traditional method to obtain the score ω_j of the alternative A_j is by means of direct assessment. At a first level a weight α_i is assigned to criterion C_i ($0 \leq \alpha_i \leq 1 \; \forall \; i = 1, \ldots, m; \; \Sigma_{i=1}^{m} \alpha_i = 1$). At a second level the alternative A_j is given a weight β_{ij} under each criterion separately ($0 \leq \beta_{ij} \leq 1 \; \forall \; i = 1, \ldots, m, \; j = 1, \ldots, n; \; \Sigma_{j=1}^{n} \beta_{ij} = 1 \; \forall \; i = 1, \ldots, m$). Then the scores of the alternatives are computed according to

$$\omega_j = \sum_{i=1}^{m} \alpha_i \beta_{ij} \qquad (j = 1, \ldots, n) \qquad (A1)$$

In practice, however, the weights α_i and β_{ij} are frequently unknown. A common approach to this problem is to obtain estimates via a method of pairwise comparison (cf. Saaty (1980)). At the first level each of the N committee members is asked if he or she prefers either the k-th criterion or the ℓ-th for all $k > \ell$ ($k, \ell = 1, \ldots, m$). Further it is assumed that the probability of preferring criterion C_k over C_ℓ is equal to $p_{k\ell} = \alpha_k/(\alpha_k + \alpha_\ell)$. Then the probability that $N_{k\ell}$ of N independent committee members will vote for C_k rather than C_ℓ is given by the binomial probability distribution

$$\binom{N}{N_{k\ell}} \left(\frac{\alpha_k}{\alpha_k + \alpha_\ell}\right)^{N_{k\ell}} \left(\frac{\alpha_\ell}{\alpha_k + \alpha_\ell}\right)^{N - N_{k\ell}} \qquad (A2)$$

Hence, assuming that the comparisons of different pairs of criteria are independent as well, the likelihood function of the unknown parameters $\alpha_1, \ldots, \alpha_m$ is equal to

$$L(\alpha_1, \ldots, \alpha_m) = \prod_{k=2}^{m} \prod_{\ell=1}^{k-1} \binom{N}{N_{k\ell}} \left(\frac{\alpha_k}{\alpha_k + \alpha_\ell}\right)^{N_{k\ell}} \left(\frac{\alpha_\ell}{\alpha_k + \alpha_\ell}\right)^{N-N_{k\ell}} \qquad (A3)$$

Bradley and Terry (1952) use (A3) to find the maximum likelihood estimates for the weights $\alpha_1, \ldots, \alpha_m$ of the criteria. At the second level the maximum likelihood estimates of the weights β_{ij} of the alternatives for each criterion separately are determined in a similar way. Substitution of the computed estimates in (A1) then yields the scores for each of the alternatives.

In this appendix we introduce a Bayesian analysis of (A3), where the prior distributions of the α_i and β_{ij} are assumed to be uniform, and we apply the mixed integration method to compute the posterior expected values and covariances of these unknown parameters. Then one can determine the posterior expected values and standard deviations of the final weights $\omega_1, \ldots, \omega_n$, using (A1) and statistical results with respect to the expected value and variance of sums and products of random variables.

Next, we compare the results of the maximum likelihood and the Bayesian approach on the example mentioned above. The data for the problem are presented in Table 1. For each pair of factors C_i/C_j and A_i/A_j we depicted the number of committee members which preferred the i-th factor over the j-th, as well as the total number of voters. Note that not all members expressed their opinion on each pair of factors. Furthermore, if a committee member considered the i-th and j-th factor equally important this vote is attributed half to each of these factors.

Table 1. Data for the Operations Research professorship

C_2/C_1	C_3/C_1	C_3/C_2	C_4/C_1	C_4/C_2	C_4/C_3
(1.5, 3)	(.5, 1)	(0, 2)	(3, 3)	(1, 3)	(1, 1)

	A_2/A_1	A_3/A_1	A_3/A_2
C_1	(1, 2)	(1.5, 2)	(1, 1)
C_2	(0, 1)	(0, 1)	(0, 0)
C_3	(0, 3)	(0, 1)	(.5, 1)
C_4	(0, 0)	(1, 2)	(0, 1)

Given these pairwise comparisons the likelihood functions are given by (A4) up to (A8), apart from normalizing constants. Note that since we have chosen our prior distributions uniform the kernels of the posterior distributions are proportional to the likelihood functions.

$$L_1(\alpha_1,\alpha_2,\alpha_3,\alpha_4) \propto \left(\frac{\alpha_2}{\alpha_1+\alpha_2}\right)^{1\frac{1}{2}}\left(\frac{\alpha_1}{\alpha_1+\alpha_2}\right)^{1\frac{1}{2}}\left(\frac{\alpha_3}{\alpha_1+\alpha_3}\right)^{\frac{1}{2}}\left(\frac{\alpha_1}{\alpha_1+\alpha_3}\right)^{\frac{1}{2}} * \qquad \text{(A4)}$$

$$* \left(\frac{\alpha_2}{\alpha_2+\alpha_3}\right)^{2}\left(\frac{\alpha_4}{\alpha_1+\alpha_4}\right)^{3}\left(\frac{\alpha_4}{\alpha_2+\alpha_4}\right)^{1}\left(\frac{\alpha_2}{\alpha_2+\alpha_4}\right)^{2}\left(\frac{\alpha_4}{\alpha_3+\alpha_4}\right)^{1},$$

(A5)

$$L_2(\beta_{11},\beta_{12},\beta_{13}) \propto \left(\frac{\beta_{12}}{\beta_{11}+\beta_{12}}\right)^{1}\left(\frac{\beta_{11}}{\beta_{11}+\beta_{12}}\right)^{1}\left(\frac{\beta_{13}}{\beta_{11}+\beta_{13}}\right)^{1\frac{1}{2}}\left(\frac{\beta_{11}}{\beta_{11}+\beta_{13}}\right)^{\frac{1}{2}}\left(\frac{\beta_{13}}{\beta_{12}+\beta_{13}}\right)^{1},$$

$$L_3(\beta_{21},\beta_{22},\beta_{23}) \propto \left(\frac{\beta_{21}}{\beta_{21}+\beta_{22}}\right)^{1}\left(\frac{\beta_{21}}{\beta_{21}+\beta_{23}}\right)^{1}, \qquad \text{(A6)}$$

$$L_4(\beta_{31},\beta_{32},\beta_{33}) \propto \left(\frac{\beta_{31}}{\beta_{31}+\beta_{32}}\right)^{3}\left(\frac{\beta_{31}}{\beta_{31}+\beta_{33}}\right)^{1}\left(\frac{\beta_{32}}{\beta_{32}+\beta_{33}}\right)^{\frac{1}{2}}\left(\frac{\beta_{33}}{\beta_{32}+\beta_{33}}\right)^{\frac{1}{2}}, \qquad \text{(A7)}$$

$$L_5(\beta_{41},\beta_{42},\beta_{43}) \propto \left(\frac{\beta_{41}}{\beta_{41}+\beta_{43}}\right)^{1}\left(\frac{\beta_{43}}{\beta_{41}+\beta_{43}}\right)^{1}\left(\frac{\beta_{42}}{\beta_{42}+\beta_{43}}\right)^{1}. \qquad \text{(A8)}$$

The resulting maximum likelihood estimates, which coincide with the posterior modes, have been determined by a constrained numerical optimization routine: we used the TWOFAS-package of Louter and Van der Hoek (1984). In order to compute the Bayesian estimates we used the mixed integration method. Our results are shown in Table 2. It is of considerable interest to observe that candidate A_1 would definitely prefer a Bayesian procedure for choosing the new professor. However, further inspection of the posterior standard deviations indicates that the difference is not as clear-cut as suggested by the maximum likelihood estimates.

Table 2. Results for the Operations Research professorship

Maximum likelihood

α_1	α_2	α_3	α_4
.111	.384	.037	.466

β_{ij}	1	2	3
1	.158	.119	.723
2	1.000	.000	.000
3	1.000	.000	.000
4	.000	1.000	.000

	(.187)	(.118)	(.164)
4	.304	.454	.242
	(.195)	(.228)	(.167)

ω_1	ω_2	ω_3
.439	.479	.080

Bayes

α_1	α_2	α_3	α_4
.152	.357	.106	.385
(.088)	(.141)	(.090)	(.148)

β_{ij}	1	2	3
1	.256	.239	.505
	(.152)	(.158)	(.195)
2	.534	.233	.233
	(.210)	(.184)	(.184)
3	.623	.154	.223

ω_1	ω_2	ω_3
.413	.311	.277
(.127)	(.128)	(.110)

APPENDIX B

The set Ω in equation (8) may be defined as

$$\Omega = \{\eta | \eta \in \mathbb{R}^{s-1} \text{ and } \eta' V_1 \eta \leq 1\} \qquad \text{B1}$$

where

$$V_1 := P - q \frac{1}{r} q' \qquad \text{B2}$$

The characterization of Ω is based on the following argument. We start with (2) and (3) and note that v is a real number. Next, given an (η, ρ) the corresponding value of v is obtained from the inverse transformation (9). Suppose the discriminant $D(\eta)$ in (9) is negative. Then there is no real solution for v. This is a contradiction. Therefore, using the definition of $D(\eta)$ in (10), we have

$$\eta' q q' \eta - r \eta' P \eta + r \geq 0 \qquad \text{B3}$$

This inequality can be rewritten, by making use of B2), as

$$\eta' V_1 \eta \leq 1 \qquad \eta' V_1 \eta \leq 1 \qquad \text{B4}$$

So Ω is an ellipsoid provided V_1 is positive definite. It is straightforward to verify, using (2) and (3), that V_1 is the inverse of the covariance matrix of the vector u. The covariance matrix of the multivariate normal random variable θ in (2) is positive definite. Then the covariance matrix (and the inverse of the covariance matrix) of u is positive definite [cf. Anderson (1958), p. 337].

REFERENCES

Anderson, T.W., 1958, An introduction to multivariate statistical analysis (John Wiley, New York).

Bradley, R.A. and M.E. Terry, 1952, The rank analysis of incomplete block designs. I. The method of paired comparisons, Biometrika 39, 324-345.

Hammersley, J.M. and D.C. Handscomb, 1964, Monte Carlo methods (Methuen, London).

Kendall, M.G. and A. Stuart, 1969, The advanced theory of statistics (Vol. 1, third edition) (Charles Griffin, London).

Kloek, T. and H.K. van Dijk, 1978, Bayesian estimates of equation system parameters, an application of integration by Monte Carlo, Econometrica 46, 1-19. Reprinted in: A. Zellner ed., 1980, Bayesian analysis in Econometrics and Statistics (North-Holland, Amsterdam).

Lootsma, F.A., 1980, Saaty's priority theory and the nomination of a senior professor in operations research, European Journal of Operations Research 4, 380-388.

Louter, A.S. and G. van der Hoek, 1984, TWOFAS: A 2-phase code for constrained nonlinear programming, Report 8411/I, Econometric Institute, Erasmus University Rotterdam, The Netherlands.

Rao, C.R., 1973, Linear statistical inference and its applications (second edition)(John Wiley, New York).

Rubinstein, R.Y., 1981, Simulation and the Monte Carlo method (John Wiley, New York).

Trotter, H.F. and J.W. Tukey, 1956, Conditional Monte Carlo for normal samples, in: H.A. Meyer ed., Symposium on Monte Carlo methods (John Wiley, New York).

Van Dijk, H.K. and T. Kloek, 1980, Further experience in Bayesian analysis using Monte Carlo integration, Journal of Econometrics 14, 307-328.

Van Dijk, H.K. and T. Kloek, 1982, Posterior moments of the Klein Goldberger model, Report 8221/E, Econometric Institute, Erasmus University Rotterdam, to appear in: P.K. Goel and A. Zellner, eds., Bayesian inference and decision techniques with applications (North Holland, Amsterdam).

Van Dijk, H.K. and T. Kloek, 1983a, Monte Carlo analysis of skew posterior distributions: an illustrative econometric example, The Statistician 32, No. 1 and 2, 216-223.

Van Dijk, H.K. and T. Kloek, 1983b, Experiments with some alternatives for simple importance sampling in Monte Carlo integration, Report 8326/E,

Econometric Institute, Erasmus University Rotterdam, to appear in: Proceedings of the second international meeting on Bayesian Statistics, Sept. 6-10, Valencia, Spain.

Saaty, T.L., 1980, The analytical hierarchy process (McGraw-Hill, New York).

LIST OF REPORTS 1985

8500 "Publications of the Econometric Institute Second Half 1984: List of Reprints 378-400, Abstracts of Reports".

8501/O R.M. Karp, J.K. Lenstra, C.J.H. McDiarmid and A.H.G. Rinnooy Kan, "Probabilistic analysis of combinatorial algorithms: an annotated bibliography", pages.

8502/E M.E. Homan, "Verschillen in consumptie tussen één en tweekostwinnerhuishoudens: een eerste analyse", pages.

8503/O A.W.J. Kolen, "The round-trip p-center and covering problem on a tree", 17 pages.

8504/O H.C.P. Berbee, C.G.E. Boender, A.H.G. Rinnooy Kan, C.L. Scheffer, R.L. Smith and J. Telgen, "Hit-and-run algorithms for the identification of nonredundant linear inequalities", 32 pages.

8505/S H. Brozius and L. de Haan, "On limiting laws for the convex hull of a sample", 10 pages.

8506/O J.B.G. Frenk and A.H.G. Rinnooy Kan, "On the rate of convergence, to optimality of the LPT rule - postscript", 5 pages.

8507/E P. Kooiman, H.K. van Dijk and A.R. Thurik, "Likelihood diagnostics and Bayesian analysis of a micro-economic disequilibrium model for retail services", 35 pages.

8508/O C.G.E. Boender, A.H.G. Rinnooy Kan and J.R. de Wit, "A Bayesian procedure for the (s,Q) inventory problem", 26 pages.

8509/E B.M.S. van Praag, S. Dubnoff and N.L. van der Sar, "From judgments to norms: measuring the social meaning of income, age and educaton", 35 pages.

8510/S B. Bode, J. Koerts and A.R. Thurik, "On shopkeepers' pricing behaviour", 19 pages.

8511/M H. Bart, I. Gohberg and M.A. Kaashoek, "Exponentially dichotomous operators and inverse Fourier transforms", 70 pages.

8512/M J. Brinkhuis, "Normal integral bases and embedding of fields", 13 pages.

8513/E P.M.C. de Boer, "On the relationship between Revankar's and Lu-Fletcher's production function", 5 pages.

8514/S L. de Haan, "Extremes in higher dimensions: the model and some statistics", 15 pages.

8515 Publications of the Econometric Intitute First Half 1985: List of Reprints 401-414, Abstracts of Reports.

8516/A A.R. Thurik and A. Kleijweg, "Cyclical effects in retail labour productivity", 20 pages.

8517/C B.M.S. van Praag and J. van Weeren, "The impact of past experiences and anticipated future on individual income judgements", 24 pages.

8518/A A.R. Thurik and J. Koerts, "Behaviour of retail entrepreneurs". 16 pages.

8519/B M. Hazewinkel, J.F. Kaashoek and B. Leynse, "Pattern formation for a one dimensional evolution equation based on Thom's River Basin Model". 24 pages.

8520/A H.K. van Dijk, T. Kloek and C.G.E. Boender, "Posterior moments computed by mixed integration", 25 pages.

Until report 8515 this series was devided in E(conometrics), S(tatistics), M(athematics) and O(perations Research). From report 8516 on it will be

A. Economics, Econometrics and Operations Research

B. Mathematics

C. Miscellaneous